Renewable and Sustainable Energy Reviews 39 (2014) 1024–1034

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects

Christian A. Gueymard

Solar Consulting Services, P.O. Box 392, Colebrook, NH 03576, USA

A R T I C L E I N F O

Article history:
Received 28 December 2013
Accepted 12 July 2014
Available online 9 August 2014

Keywords:
Direct normal irradiance
Statistical indicators
Uncertainty
Bankability
Validation

A B S T R A C T

In the context of the current rapid development of large-scale solar power projects, the accuracy of the modeled radiation datasets regularly used by many different interest groups is of the utmost importance. This process requires careful validation, normally against high-quality measurements. Some guidelines for a successful validation are reviewed here, not just from the standpoint of solar scientists but also of non-experts with limited knowledge of radiometry or solar radiation modeling. Hence, validation results and performance metrics are reported as comprehensively as possible. The relationship between a desirable lower uncertainty in solar radiation data, lower financial risks, and ultimately better bankability of large-scale solar projects is discussed.

A description and discussion of the performance indicators that can or should be used in the radiation model validation studies are developed here. Whereas most indicators are summary statistics that attempt to synthesize the overall performance of a model with only one number, the practical interest of more elaborate metrics, particularly those derived from the Kolmogorov–Smirnov test, is discussed. Moreover, the important potential of visual indicators is also demonstrated. An example of application provides a complete performance analysis of the predictions of clear-sky direct normal irradiance obtained with six models of the literature at Tamanrasset, Algeria, where high-turbidity conditions are frequent.

Contents

E-mail address: Chris@SolarConsultingServices.com

http://dx.doi.org/10.1016/j.rser.2014.07.117
1364-0321/

1. Introduction

The rapid development and deployment of solar energy technologies over the world requires considerable investments, financial risk analyses, and well-balanced policy decisions about where which technology should be deployed in priority, for instance. Ultimately, sound engineering tools and instruments are needed to deal with the uncertainties associated with the inherent variability of the solar resource and the difficulties of its precise quantitative evaluation. A wide variety of interest groups (hereafter "stakeholders") needs to know how much solar power can be generated by any installation. In most cases, these future projections must be qualified with some probability and uncertainty, particularly to evaluate the financial risks and overall "bankability" of large projects in particular. From a different perspective, experts and solar radiation scientists may require elaborate statistics to better understand the weak points of a model and where to concentrate efforts to improve it, for instance. Conversely, for the majority of stakeholders who have only limited expertise in solar radiation data, any uncertainty analysis of such data must be presented in a comprehensive form to them so that they can easily apply this information to their normal tasks.

Since the energy produced by any solar system is a direct and strong function of the incident irradiance, it is obvious that the uncertainty in the energy produced by the system over time, and hence the uncertainty in the optimal system's design parameters, is also a strong function of the uncertainty in the incident irradiance, which is itself the result of all modeling or measurement errors. The direct link between uncertainties in the solar resource and in the design and energy production of photovoltaic (PV) or concentrating solar technologies has been addressed recently [1–4].

It is well known that the aggregation of modeled results over an increasingly longer period tends to decrease their random error (see, e.g., [5,6]). It is therefore important to compare the performance of models over an appropriate averaging period. Whereas hourly and monthly radiation data have been the dominant standards for the simulation and design of solar energy systems, respectively, other needs or possibilities have surfaced in recent years. First, many radiometric stations now report data with a much shorter time step of 1- to 10-min. This allows the validation of models at such a higher frequency, which is desirable because high-frequency irradiance data are necessary for the simulation of non-linear solar systems under rapidly changing conditions, for instance. Second, at the other end of the time scale, the characterization of the solar resource over a given area is reported in terms of its mean *annual* irradiation. This is a key factor to evaluate the expected long-term energy production of a solar system, and an essential input to the financial calculations that are involved to determine the project's bankability. Cebecauer et al. [7] gave specific indications about the sources of uncertainty in contemporary methods used to derive irradiance from satellite imagery. Vignola et al. [8] discussed the different ways of obtaining bankable radiation data, and the limitations of popular products such as Typical Meteorological Years (TMYs) and of satellite-derived data. Using practical examples, Schnitzer et al. [9] showed how the lowering of the solar resource's uncertainty significantly reduces the financial risks and increases the bankability of a project. An important step to obtain a reduction in resource uncertainty is to optimally combine short-term on-site irradiance measurements and long-term satellite-derived modeled data, so as to remove as much bias as possible in the latter and obtain the desired "best estimate" [10–12]. In this context, Meyer et al. [13] addressed the issue of calculating the resulting uncertainties when irradiance data from different sources are merged. More generally, the role of reducing the irradiance data uncertainty, and the proper way to take this uncertainty into account to evaluate the bankability of concentrating photovoltaic (CPV) projects has been recently described by Leloux et al. [3].

Considering the paucity of solar radiation measurements where they would be needed for most large-scale developments, general studies and specific projects have to rely on modeled datasets at one point or another. Some essential questions immediately follow: How do these modeled datasets differ from the truth, or at least from high-quality data that would be measured locally? With what confidence level can we trust such data? How does the dataset obtained with model A compare to that from model B? etc.

The need for a better understanding of all the issues related to the validation of solar radiation datasets led to the adoption of international scientific cooperation projects. In recent years, various initiatives examined these issues from the following various angles, most notably:

- Task 36 on Solar Resource Knowledge Management of the International Energy Agency's Solar Heating & Cooling Program (IEA-SHC; http://archive.iea-shc.org/task36/), as presented in [14,15]
- Task 46 on Solar Resource Assessment and Forecasting of IEA-SHC (http://archive.iea-shc.org/task46/), a continuation of Task 36 (now ended) for the most part.
- The European MESoR project (Management and Exploitation of Solar Resource Knowledge; http://www.mesor.org), as described in various reports [16–18].
- Task V on Solar Resource Knowledge Management of SolarPACES (http://www.solarpaces.org/Tasks/Task5/task_V.htm), as described in the organization's 2011 report [19]. That task led to the description of methodological issues pertaining to the validation of direct irradiance datasets [20].

These activities have generated a lot of interest among the solar radiation community, thus creating a wide-reaching forum for fruitful dialog and methodological advances, on which some of the concepts presented here are based.

2. Historical developments and recent advances

A survey of the historical literature demonstrates that the issue of assessing the accuracy of solar radiation data and their underlying models has not attracted the attention it deserves, at least until recently. For decades, solar radiation data or model outputs have been only validated with simple conventional statistical indicators such as the mean bias difference (MBD), root mean square difference (RMSD), coefficient of determination (R^2), or (more rarely), mean absolute difference (MAD). (Formal definitions of these and other indicators appear in Section 5, as well as a discussion of the frequent

usage of the term "error" in lieu of "difference".) Additionally, qualitative results were generally obtained in the form of scatterplots comparing the predicted and measured values individually. A critique of the typical use (or overuse) of some of these statistics [21,22] has apparently not changed their prevalence in the current solar literature. A series of publications by Willmott [23–25] developed the case for the adoption of his "index of agreement", WIA, but it remained largely ignored by the solar community, with some exceptions (e.g., [26]). A revised definition of WIA was later proposed [27], but was immediately critiqued by Legates and McCabe [28], who instead proposed their "coefficient of efficiency" derived from previous work [29,30].

Stone [31,32] proposed the *t*-statistic, hereafter TS, calculated as a combination of MBD and RMSD, as a more robust indicator, and better able to facilitate the overall ranking of modeled results obtained for diverse locations. Muneer et al. [33] reviewed the existing performance indicators and suggested a new one, the "accuracy score", AS, obtained as a linear combination of 6 conventional indicators having identical weights. The selection of these indicators appears rather arbitrary, since the authors did not justify the rationale behind it, nor did they explain whether each of the 6 selected indicators would carry a relatively similar information per unit AS. A drawback of AS, as previously noted [34], and the reason why it will not be discussed further, is that it is not an absolute metric: its value changes depending on which models are being compared, or their number.

Gueymard and Myers [34] offered detailed guidelines about how to properly conduct a validation exercise, and compared the merits and weaknesses of various indicators, including WIA, TS and AS, with respect to a case study, consisting of a limited validation of clear-sky irradiance values obtained with various models. Other discussions on the relative merits of various statistical indicators for solar radiation model evaluation have appeared in publications pertaining to various disciplines (e.g., [35–37]).

From a different perspective, the development and availability of many extensive solar radiation datasets, most generally derived from satellite observations but all using different methods and models, has prompted renewed interest for the issue of their validation. Moreover, the need for more detailed scrutiny emerged. For instance, many solar thermal systems, particularly those relying on concentrators and usually referred to as "concentrating solar power" (CSP), are inherently non-linear. By design they usually cannot work well or at all under low-irradiance conditions. Depending on technology, the threshold irradiance may vary (more or less dynamically due to the system's inertia) between 100 and 500 W/m^2 [38]. The accuracy of predictions below a specific CSP system's threshold is therefore of virtually no interest for its stakeholders. Conversely, the accuracy of predicted irradiance around the design point of any CSP system is of utmost importance. Since this design point is typically project-specific, it becomes necessary to evaluate the accuracy of the irradiance predictions over a sizeable range of values that encompass the design point. This can be achieved by analyzing the relative frequency distributions of the predictions and reference (measured) data. Using such frequency distributions and probabilistic modeling, Ho et al. [2] identified that, by a considerable margin, the major source of uncertainty in the simulation of a CSP system (of the central tower type with thermal storage in their case study) was the solar resource (direct irradiation in their case). This makes the determination of the solar radiation data uncertainty all the more important for such systems.

Moreover, the use of TMYs is popular for the energy simulation of solar energy systems or buildings. The constructions of TMYs also assume a good frequency distribution of the solar radiation variables. Because of its practical importance, the use of frequency distributions to assess datasets quality will be discussed further in Section 5.3.

In parallel, the production of high-quality solar radiation *forecasts* (as opposed to the historical data mostly discussed so far) is now becoming a strong research goal, due to the increasing penetration of variable sources of electricity production (solar and wind) in electric grids. A legitimate question that is now hotly debated between experts is which statistical indicators should be used to assess the performance of solar forecasts. Whereas Hoff et al. [36] recommend the use of MAD based on qualitative reasoning, Marquez and Coimbra [39] propose a new performance metric to evaluate how a forecast model is able to effectively predict the stochastic variability of the irradiance. Since the field of solar forecasting is in rapid evolution, it is expected that the debate about which performance metric to use will intensify. For this reason, the rest of this report will exclusively focus on the validation of historical modeled datasets.

3. Terminology and scope

To document the quality of modeled data, terms such as "validation", "evaluation", "verification" or "benchmarking" are used more or less interchangeably in the literature. However, based on an extensive analysis of the implications of this terminology, Oreskes et al. [40] reasoned that numerical models could not really be "validated" or "verified". To some stakeholders, the terms *validation* or v*erification* may just mean that "the model works", without any further quantification. For these reasons, the term "performance assessment" appears more appropriate when a *precise quantification* of the accuracy of modeled data is the goal. However, this semantic question can be regarded as a minor point, inasmuch as the goal of the analysis is clearly defined. Throughout this report, the term "validation" is simply used as a synonym for "performance assessment".

Another level of terminology must be added to clarify *what exactly* is being validated. One possible goal is to evaluate the performance of the radiative models themselves, i.e., their "intrinsic" performance. Examples of such studies include [41–45]. To achieve that goal, the models' input data must be of the highest quality and lowest uncertainty possible, so that the difference between the modeled results and the reference observations can be attributed almost entirely to the models, rather than to their input data. This means that the latter must be obtained with collocated instruments of the highest possible accuracy to provide "nearly-perfect" inputs, without any temporal or spatial interpolation. This is an ideal case, and for research purposes only, since these conditions are essentially never met in practical situations. In contrast, it is also possible to rather focus on the "overall performance" of the combination of radiation models and their input data, in the case the latter are of insufficient quality (e.g., because interpolated or estimated), or if the inputs need to be considered as indissociable from the model itself—which is typically the case with satellite-derived data series, for instance. This was the *de facto* approach followed in many studies, e.g., [46–64].

The accuracy of modeled data can be assessed in many different ways, depending on the type of model, temporal resolution of the data, etc. Various rules must be observed to obtain meaningful results [34], and various statistical indicators can be used, as detailed in Section 5. The role of these statistical indicators is to evaluate the performance of large modeled datasets with only a few numbers. When many modeled datasets are compared to a single reference (presumably measured irradiance data), it becomes possible to obtain their ranking. This ranking usually depends on the statistical indicators selected, and cannot therefore be considered an absolute metric [34,65].

4. Methodology

To obtain valid performance results, it is obvious that the comparison between modeled (predicted) and reference (measured) data must only include comparable data points and be commensurate. This is relatively easy to achieve when using “instantaneous” data, such as data reported with a time step of 1 min, which has become standard at research-class radiometric sites. Ideally, each institution should evaluate an instantaneous uncertainty or provide a quality flag with each data point, as done by the National Renewable Energy Laboratory (NREL) for instance [66], but this is still more the exception than the rule. In any case, it is the analyst's responsibility to perform a quality control (QC) procedure and eliminate all reference data points that are out of bounds. When dealing with the performance of models under specific conditions, such as clear skies for instance, another level of difficulty (and inadvertent source of error) is added by the filter needed to eliminate all unwanted conditions (e.g., cloudy periods). All kinds of filtering techniques have been used to eliminate cloudy periods, from a crude minimum threshold value imposed on the clearness index,[1] k_t, to a sophisticated and dynamic algorithm that involves all three components (direct, diffuse and global), which must be independently measured at a relatively high frequency [67,68].

When the goal of the analysis is to validate datasets on a longer time scale (e.g., monthly data), other difficulties arise. A major one is related to data breaks, caused by missing reference data points. No instrument or experimental setup being perfect, all long-term observation datasets are unfortunately incomplete to some degree, or at least contain egregious data, which may or may not have been flagged as such by the QC algorithm. In the absence of internationally accepted procedures to deal with this issue, each institution has a different approach. Egregious data points may be replaced by interpolated/extrapolated data, be simply eliminated completely, or just be flagged but unused for temporal averaging, etc. The way missing or egregious data points are dealt with may affect the long-term (e.g., monthly) averaging process, and hence the validation results [69]. The specific QC and averaging procedures adopted by the Baseline Surface Radiation Network (BSRN; http://www.bsrn.awi.de/) of the World Radiation Monitoring Center (WRMC) are described by Roesch et al. [70]. BSRN is the world network that maintains the highest standard of quality, which makes its data the source of choice when validating any type of modeled surface radiation data. A unification of the QC and averaging methods adopted by all institutions involved in solar radiation monitoring is desirable, but will certainly require precise guidelines from the World Meteorological Organization (WMO), to which BSRN is affiliated.

5. Statistical indicators

In what follows, the ith observed data point will be noted o_i and the ith predicted (modeled) data point will be noted p_i. The mean values of the two distributions (each totaling N points) are noted O_m and P_m, respectively. The ith modeled-measured difference in the distribution is $p_i - o_i$. This difference is customarily referred to as an “error”. This usage masks the fact that o_i itself is imperfect and contains error, and is thus incorrect. In this context, the term "prediction error" is only acceptable when it is known for sure that the reference data has a very low or negligible uncertainty compared to the modeled values.

[1] Ratio between the global horizontal irradiance and its extraterrestrial counterpart.

The possible statistical indicators (or “metrics”) that can be used in performance assessment studies will be divided into the following four categories:

- Class A: indicators of the dispersion (or “error”) of individual points; their value would be 0 for a perfect model.
- Class B: indicators of overall performance; their maximum value is 1 (for a perfect model).
- Class C: indicators of distribution similitude.
- Class D: visual (qualitative) indicators.

5.1. Class A—indicators of dispersion

These are the indicators that the majority of readers should be most familiar with. They are all expressed here in percent (of O_m) rather than in absolute units (W/m^2 for irradiances, or MJ/m^2 or kWh/m^2 for irradiations) because non-expert stakeholders can much more easily understand percent results. In any case, stating the value of O_m in all validation results allows the experts to convert back the percent figures into absolute units if they so desire. Formulas in this section are well established and do not need further references.

5.1.1. Mean bias difference (MBD)

For the reason mentioned above, it is also referred to as Mean bias error (MBE). It is obtained as

$$\mathrm{MBD} = (100/O_m)\sum_{i=1}^{i=N}(p_i - o_i) \tag{1}$$

5.1.2. Root mean square difference (RMSD)

$$\mathrm{RMSD} = (100/O_m)\left[\sum_{i=1}^{i=N}(p_i - o_i)^2/N\right]^{1/2} \tag{2}$$

5.1.3. Mean absolute difference (MAD)

$$\mathrm{MAD} = (100/O_m)\sum_{i=1}^{i=N}|p_i - o_i| \tag{3}$$

5.1.4. Standard deviation of the residual (SD)

$$\mathrm{SD} = (100/O_m)\left[\sum_{i=1}^{i=N}N(p_i - o_i)^2 - \left\langle\sum_{i=1}^{i=N}(p_i - o_i)\right\rangle^2\right]^{1/2}\Big/N \tag{4}$$

5.1.5. Coefficient of determination (R^2)

$$R^2 = \left[\left[\sum_{i=1}^{i=N}(p_i - P_m)(o_i - O_m)\right]\Big/\left[\sum_{i=1}^{i=N}(p_i - P_m)^2(o_i - O_m)^2\right]\right]^2 \tag{5}$$

5.1.6. Slope of best-fit line (SBF)

$$\mathrm{SBF} = \left[\sum_{i=1}^{i=N}(p_i - P_m)(o_i - O_m)\right]\Big/\left[\sum_{i=1}^{i=N}(o_i - O_m)^2\right] \tag{6}$$

5.1.7. Uncertainty at 95% (U_{95})

$$U_{95} = 1.96\ (\mathrm{SD}^2 + \mathrm{RMSD}^2)^{1/2} \tag{7}$$

5.1.8. t-statistic (TS)

$$\mathrm{TS} = [(N-1)\mathrm{MBD}^2/(\mathrm{RMSD}^2 - \mathrm{MBD}^2)]^{1/2} \tag{8}$$

5.2. Class B—indicators of overall performance

These are indicators that are less common in the solar field than those of Class A. They convey relatively similar information as those of Class A, with the cosmetic advantage that a higher value indicates a better model.

5.2.1. Nash–Sutcliffe's efficiency (NSE)

As defined in [30]

$$\mathrm{NSE} = 1 - \left[\sum_{i=1}^{i=N}(p_i - o_i)^2\right] \Big/ \left[\sum_{i=1}^{i=N}(o_i - O_m)^2\right] \tag{9}$$

5.2.2. Willmotts's index of agreement (WIA)

As defined in [23]

$$\mathrm{WIA} = 1 - \left[\sum_{i=1}^{i=N}(p_i - o_i)^2\right] \Big/ \left[\sum_{i=1}^{i=N}(|p_i - O_m| + |o_i - O_m|)^2\right] \tag{10}$$

5.2.3. Legates's coefficient of efficiency (LCE)

As defined in [28,29]:

$$\mathrm{LCE} = 1 - \left[\sum_{i=1}^{i=N}|p_i - o_i|\right] \Big/ \left[\sum_{i=1}^{i=N}|o_i - O_m|\right] \tag{11}$$

LCE and NSE vary between 1 for perfect agreement and $-\infty$ for complete disagreement, whereas WIA varies only between 1 and 0.

5.3. Class C—indicators of distribution similitude

Here, the goal is to compare one or more cumulative frequency distribution of modeled data to that of a reference dataset. Can one or more single number provide a measure of the similitude between two or more distributions? Substantial progress in that direction resulted from an initial study by Polo et al. [71], who proposed to use the Kolmogorov–Smirnov test when comparing different cumulative distribution functions (CDFs), because of its advantage of being non-parametric and valid for any kind of CDF. Espinar et al. [72] developed the method further, now referring to it as the Kolmogorov–Smirnov test Integral (KSI). KSI (in percent) is defined as

$$\mathrm{KSI} = \frac{100}{A_c}\int_{X_{\min}}^{X_{\max}} D_n dx, \tag{12}$$

where D_n is the absolute difference between the two normalized distributions within irradiance interval n, X_{min} and X_{max} are the minimum and maximum values of the binned reduced irradiance, x, and A_c is a characteristic quantity of the distribution

$$A_c = D_c(X_{max} - X_{min}) \tag{13}$$

where the critical value, D_c, is a statistical characteristic of the reference distribution, defined as a function of its number of points, N

$$D_c = \Phi(N)/N^{1/2}, \tag{14}$$

and $\Phi(N)$ is a pure function of N, for which an accurate numerical approximation has been obtained [73]. As N increases and tends to infinity, $\Phi(N)$ tends to its asymptotic value of 1.628. (Espinar et al. simplified this by assuming that $\Phi(N)$ is constant at ≈ 1.63.) KSI is 0 if the two distributions being compared can be considered identical in a statistical sense.

Espinar et al. also added the OVER statistic, which is derived from KSI. OVER describes the relative frequency of exceedence situations, when the normalized distribution of modeled data points in specific bins exceeds the critical limit that would make it statistically undistinguishable from the reference distribution. OVER (in percent) is obtained as

$$\mathrm{OVER} = \frac{100}{A_c}\int_{X_0}^{X_1} \mathrm{Max}(D_n - D_c, 0)dx \tag{15}$$

OVER is 0 if the normalized distribution always remains below D_c. The reader is referred to [72] for details about the calculation of KSI and OVER. Espinar et al. applied this technique to the validation of a satellite-derived global radiation dataset against 38 radiometric stations in Germany. Interestingly, the stations where MBD and RMSD were smallest were not necessarily those that had the smallest KSI or OVER, and vice versa. OVER was null at 34 of the 38 stations, indicating that the satellite-derived dataset was generally respecting the distribution of global irradiance over most of Germany. This example showed that the use of KSI and OVER brings a different kind of information than the more conventional indicators of Class A or B, and can also be more discriminant (OVER most particularly).

This author [43] improved the calculation of D_c, per Eq. (14), and proposed a Combined Performance Index (CPI) such that

$$\mathrm{CPI} = (\mathrm{KSI} + \mathrm{OVER} + 2\mathrm{RMSE})/4. \tag{16}$$

where all values are expressed in percent. The interest of CPI is that it combines conventional information about dispersion and bias (through RMSE) with information about distribution likeness (through KSI and OVER), while maintaining a high degree of discrimination between different models. The latter feature is of course highly desirable when comparing different models of similar performance. It is now argued that, if a single statistical indicator had to be selected to powerfully compare the performance of models, the best choice would be CPI.

5.4. Class D—visual indicators

This category is completely different from the three previous ones because the goal here is to obtain a visualization rather than summary statistics in the form of a few numbers.

The most widely used visual tool is certainly the scatterplot. It directly compares the predicted and the reference (measured) data. Perfect predictions would be aligned along the 1:1 diagonal. A drawback of such a plot is that it is nearly impossible to combine the results of more than two models on the same graph without losing legibility.

One visualization tool that is popular in various disciplines, but not in the solar field yet, is the Taylor diagram [74]. It combines information about RMSD, SD and R^2 into one single polar diagram. An example of such diagrams is given in Section 5.5 below. Because of the broad interest that followed the introduction of the Taylor diagram, scripts or codes have been written in various languages, such as python or R, so that the preparation of such diagrams is now relatively easy. One interest of this diagram is that many different models can be compared on a single diagram, which gives a very rapid assessment of those that are closest to the reference dataset. However, if all models perform relatively similarly overall, their representative points become too close or superimposed, and the diagram is then not discriminant enough to be useful.

A "mutual information diagram", which consists of a revision of the Taylor diagram, has been proposed recently [75], based on mutual information theory. It has potentially more discriminating power than its predecessor, but is still hampered by the current lack of scripts or software code to obtain it easily.

Another kind of diagram with high visualization potential is the boxplot, also known as box-and-whisker plot (http://en.wikipedia.org/wiki/Box_plot). Like the Taylor diagram, it is popular in many disciplines, but not that much in the solar field. One simple version of the boxplot, showing only the binned mean error and its

standard deviation as a function of a user-selected variable, has been introduced by Ineichen et al. in 1984 [76]. Ineichen, as well as this author, then used this type of plot in various later publications (e.g., [77,78]). A slightly different representation, using the binned RMSD rather than the binned SD, has also been used [79]. An advantage of this plot is that the binned error can be displayed as a function of any variable, e.g., the measured irradiance, air mass, or a significant input to the model (such as cloud fraction or aerosol optical depth) to study the sensitivity of the model output to that variable. A drawback is that it is difficult to combine more than about 3 different series of model results on the same graph without losing legibility.

5.5. Example of application: clear-sky direct irradiance predictions

Considering the discussion in Section 3, the present exercise is designed to evaluate the *intrinsic* performance of various datasets. It focuses on broadband irradiance models that use atmospheric data to predict the cloudless-sky direct normal irradiance (DNI). From the large inventory of such models that have been proposed in the literature, six of them have been selected here for their different types of input variables, from simple to detailed.

The Meinel and Meinel model [80] is the simplest of the group, since—besides solar zenith angle, Z, which is an input to all models—it only depends on site's elevation, h (in km) according to:

$$E_{bnM} = E_{sc}[0.14h+(1-0.14h)\exp(-0.357\cos^{-0.678}Z] \qquad (17)$$

where E_{sc} is the solar constant. This model is recommended by an online study reference for PV applications (http://pveducation.org/pvcdrom), and was recently selected to evaluate the theoretical performance of a novel concentrating photovoltaic thermal (CPV/T) system for building applications [81].

The Allen model [82–84] consists of a modification of an older model [85], which was recommended in an early handbook on solar energy [86]. The Allen model is a function of site pressure, p (in kPa), and precipitable water, w (in cm), through

$$E_{bnA} = 0.98S\,E_{sc}\exp[-0.00146\,(p/\cos Z)-0.162(w/\cos Z)^{0.25}] \qquad (18)$$

where S is the sun–earth distance correction factor.

The Ineichen–Perez model [59,87] is a function of h and the Linke turbidity coefficient, T_L, which takes the extinction effects of aerosols and water vapor into account. The DNI formulations in the two references just mentioned are slightly different. The version of reference [59] is considered the official version by its authors (Pers. Comm. with Richard Perez, 2013), since it was used to derive gridded irradiance estimates using cloud radiance data from satellites in combination with an early version of the SUNY radiation model:

$$E_{bnIP} = SE_{sc}(0.664+0.163\exp(h/8)\exp[-0.09m\,(T_L-1)] \qquad (19)$$

An alternate expression is used when $T_L < 2$ [87].

The Ineichen model [88] is more elaborate, since it uses p, w and τ_{a700} as inputs, where τ_{a700} is the aerosol optical depth (AOD) at 700 nm. Among other applications, this model was selected to obtain clear-sky irradiances in the current version of the SUNY satellite model, thus replacing the model mentioned just above. The governing equations are fully described in the original

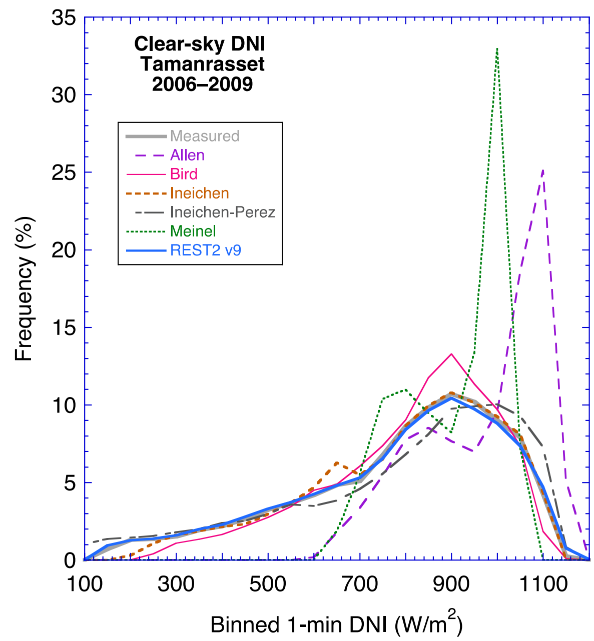

Fig. 1. Frequency distribution of DNI as measured and predicted by the six models at Tamanrasset.

Table 1
Performance statistics of six clear-sky radiation models tested against 1-min measured data at Tamanrasset, Algeria. The percent statistics refer to the mean observed DNI, 761.6 W/m^2 for N=69,710. The inputs required by each model for the calculation of DNI are also indicated. Based on the number of inputs, the simplest models appear in the leftmost columns, whereas the most detailed models appear in the rightmost columns.

Model	Meinel	Allen	Ineichen–Perez	Ineichen	Bird	REST2 v9
Inputs (for DNI)	h	p, w	h, T_L	p, w, τ_{a700}	p, w, u_o, α,β	p, w, u_o, u_n, α,β
Mean DNI (W/m^2)	879.3	947.0	764.5	771.6	784.4	757.6
Class A (%)						
MBD	15.5	24.3	0.4	1.3	3.0	−0.5
RMSD	28.2	32.8	6.0	5.4	6.7	2.2
MAD	20.3	25.0	4.9	2.9	4.8	1.5
SD	23.6	22.0	6.0	5.3	6.0	2.1
R^2	0.3754	0.4500	0.9769	0.9708	0.9864	0.9946
SBF	0.3027	0.3995	1.1141	0.9119	0.8193	1.0042
U95	55.2	64.3	11.8	10.7	13.1	4.3
TS	173.3	292.5	16.7	66.2	132.6	65.4
Class B						
NSE	0.0867	−0.2371	0.9582	0.9660	0.9487	0.9945
WIA	0.6358	0.6406	0.9907	0.9908	0.9846	0.9986
LSE	0.1477	−0.0480	0.7927	0.8772	0.7980	0.9352
Class C (%)						
KSI	1809.7	2592.2	370.3	153.9	465.1	79.8
OVER	1725.0	2506.2	283.2	95.3	381.5	14.9
CPI	897.8	1291.0	166.4	65.1	215.0	24.9

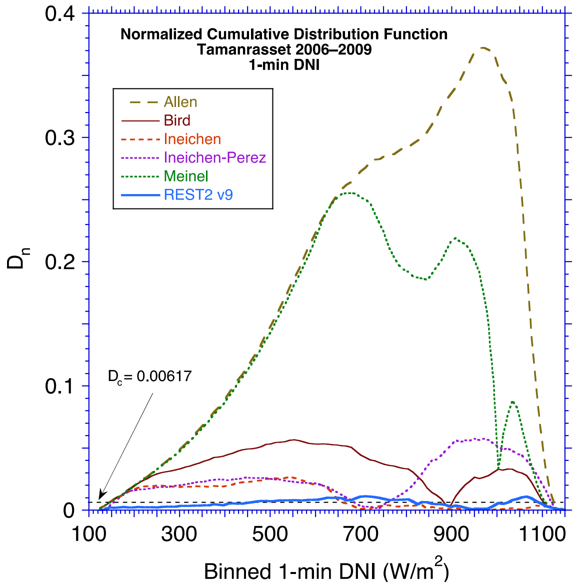

Fig. 2. Absolute difference between the measured and modeled normalized distributions, using the same six models as in Fig. 1 and DNI data from Tamanrasset.

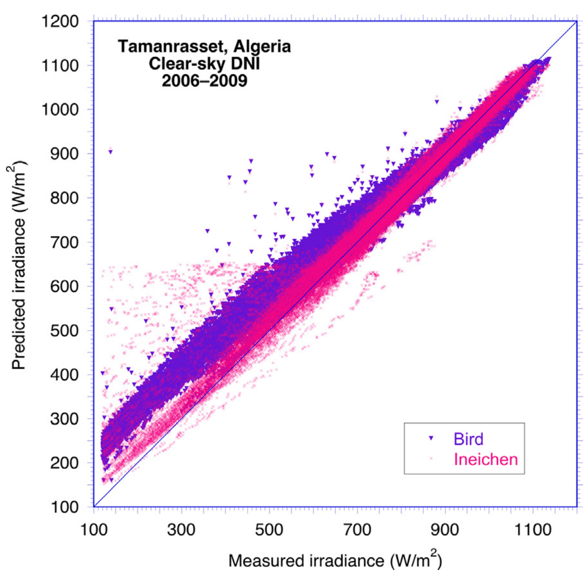

Fig. 3. Scatterplot of DNI predictions with the Bird and Ineichen models compared to measured data at Tamanrasset.

reference [88]. Values of τ_{a700} are obtained here according to the method described in [43].

The Bird model [89,90] is one of the earliest broadband transmittance models, where each major atmospheric extinction process is described by a specific transmittance function. In addition to p and w, the model's DNI calculations also depend on the ozone amount, u_o, and the Ångström turbidity coefficients α and β, per the implementation explained in [43,91].

Finally, version 9 of REST2 [92] represents the case of a more elaborate model, due to its two-band formulation and its longer list of inputs. Those inputs pertaining to the prediction of DNI include p, w, u_o, α, β, and the total NO_2 amount, u_n.

The six models described above are tested here against 1-min DNI observations from the Tamanrasset, Algeria BSRN station (latitude 22.790°N, longitude 5.529°E, elevation 1385 m). As discussed in Section 4, BSRN observation methods and quality control procedures are sophisticated and guarantee high-quality data as a result. Because of its relatively high elevation, Tamanrasset often experiences very low water vapor and aerosol optical depths. However, dust storms are frequent in the region and do increase AOD considerably on a more or less regular basis. This makes AOD's magnitude largely variable, and very high at times. DNI is measured there with an Eppley NIP pyrheliometer, whereas precipitable water and AOD are measured with a Cimel CE-318 sunphotometer from a collocated AERONET station. The highest quality data ("Level 2") are used for AOD and precipitable water. The same methodology as described in [43] is used here to prepare the input data and the coincident data points (i.e., DNI data within $\pm$ 1.5 min from sunphotometer data, and measured DNI > 120 W/m^2), except that the requirement of a completely cloudless sky is relaxed here, since only a clear line-of-sight to the sun is needed to obtain valid measurements for both DNI and AOD.

Data for the common observation period 2006–2009 are used here, and provide N=69,710 valid data points, whose mean measured DNI is 761.6 W/m^2. Table 1 provides the performance results using all indicators from Class A, B and C. As could be expected, the two models that do not use aerosol information (Allen and Meinel) do not perform as well as the other four, by a wide margin. Using indicators from Class A or B, the differences between the four other models are not as obvious. A surprising finding is that the TS results are not completely consistent with those obtained with the other indicators of Class A or B. From the denominator in Eq. (8), it appears that whenever MBD and RMSD are both low, TS becomes artificially high. Hence, TS might not be the ideal statistic for model comparison it was meant to be.

The inter-model differences noted above increase dramatically when indicators of Class C are considered.. Interestingly, none of the six models succeeds to maintain their normalized distribution below the value of D_c for that distribution. This is caused by the very low value of D_c when N is very large, as a consequence of Eq. (14).

Fig. 1 shows the frequency distributions of the measured DNI values and all modeled values. These are transformed into CDFs, which are then normalized with Eqs. (12)–(14) above. These distributions are shown in Fig. 2, along with the (very low) value of D_c from Eq. (14). As expected, the Allen and Meinel distributions are farthest from ideal. Surprisingly, however, the former is at a much larger distance than the latter, which was not expected since Allen's model has a slightly more complete description of atmospheric conditions (through w) than Meinel's. Similarly, the Bird model's distribution is relatively far from REST2's, even though the two models ultimately rely on the same aerosol inputs.

A scatterplot comparing the results of two models that use AOD information as input (Bird and Ineichen) is shown in Fig. 3. These two model exhibit different response overall, particularly under low irradiance conditions, i.e., under high air mass and/or high AOD.

A Taylor diagram providing information about the relative performances of the 6 models appears in Fig. 4. Like in Fig. 2, and in agreement with the results in Table 1, the Allen and Meinel models are quite distant from the four other models. This distance is a measure of how much information is lost by not using any AOD input to predict DNI.

Finally, a boxplot representation of the binned errors in DNI for the three models that have the lowest CPI (from Table 1) is shown in Figs. 5 and 6. The independent variable selected here is β, because it has a large impact on DNI (e.g., [93]). Conditions of very low to moderately high turbidity ($0 < \beta < 0.48$) are represented in Fig. 5. This range of values can be expected over temperate climates. Under such conditions, the Ineichen and REST2 models perform well and quite similarly. In contrast, the Ineichen–Perez model shows a strong trend of overestimating at low β and underestimating for β larger than ≈ 0.13. A different situation becomes apparent in Fig. 6 when the range of β values is extended

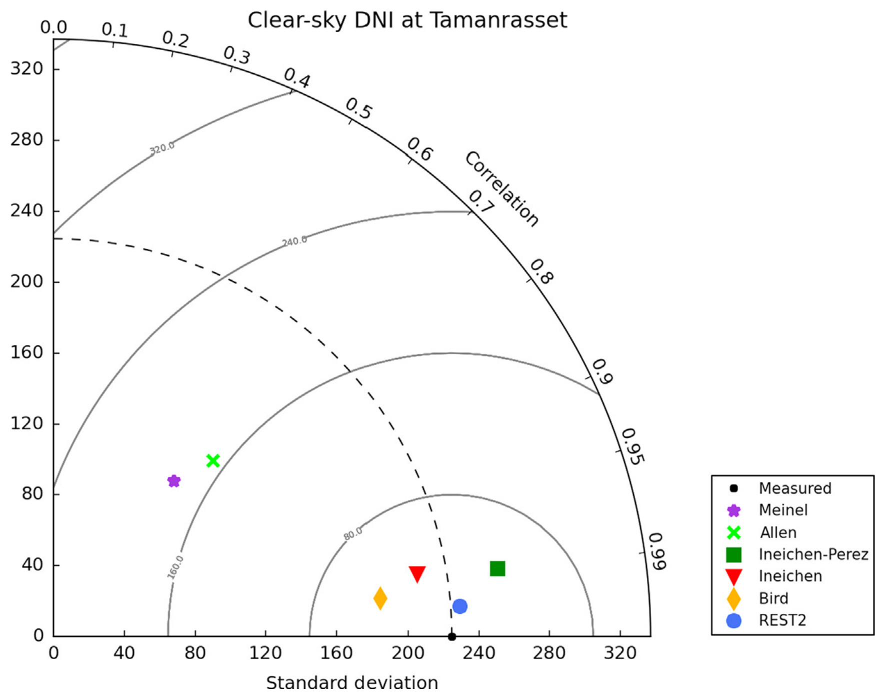

Fig. 4. Taylor diagram for the results of the six models at Tamanrasset.

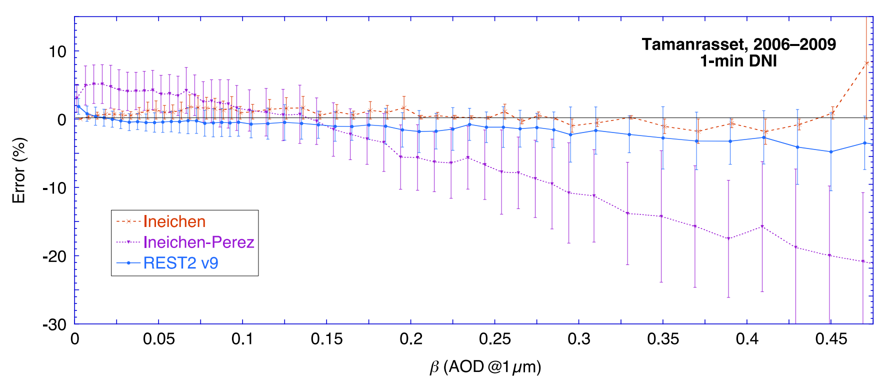

Fig. 5. Box plot showing the apparent modeled error (± 1 standard deviation) as a function of aerosol load (ß coefficient) for the three best models with the best CPI score in Table 1. The aerosol domain is limited here to $ß < 0.48$.

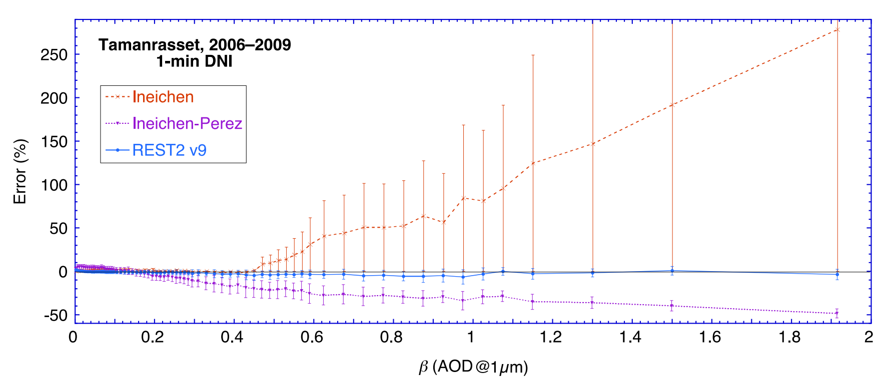

Fig. 6. Same as Fig. 5, but for unlimited ß. Note the different Y-axis scale compared to Fig. 5.

to the maximum observed (≈ 2.0) at Tamanrasset and much larger deviations become apparent. Whereas the Ineichen–Perez underprediction trend continues to intensify, the Ineichen model abruptly starts to overpredict for $ß > 0.45$, with a steep upward trend and considerable randomness (large SD), which explains the scatter for DNI < 650 W/m^2 apparent in Fig. 3. The model's abrupt change of behavior can be explained by the limited range of AOD ($0 < \tau_{a700} < 0.45$) that was used in its development [88]. High-turbidity situations with $ß > 0.45$ are generally not frequent, but still can affect the performance of solar systems, and thus should be modeled correctly. In comparison, the REST2 model keeps a low error over the whole range of ß values experienced at

Tamanrasset. The interest of Figs. 5 and 6 is to provide qualitative information on various aspects of the models' performance that otherwise would not be revealed solely by any quantitative indicator of Classes A–C or by the Taylor diagram. It is worth reminding users of radiation models that these are not supposed to be used outside of their validity range. Although this statement may appear obvious, a frequent problem is that this validity range is often *not* specified by the model's authors and therefore remains unknown to users. This is the case here for the Meinel, Allen, Ineichen–Perez and Bird models, for instance. Even when the limits are clearly indicated (as with [88] or [92] here), the magnitude of the errors that result from using the model in "out of range" mode is unknown until specific validation is performed. This often leads to the model being extrapolated outside of its specified limits under the user's assumption that it should remain "relatively" accurate anyway. This assumption is not necessarily true, and may thus lead to substantial errors in the predicted datasets, as exemplified by the results shown in Fig. 6. The accuracy of clear-sky irradiance predictions under high-turbidity conditions may have far-reaching indirect effects on the separation of the direct and diffuse components from all-sky global irradiance data [94]. Considering the strong solar developments over regions where more or less frequent high-turbidity conditions exist, users of radiative models—including institutional or commercial data providers who offer satellite-derived modeled radiation databases—should pay more attention to this off-limits accuracy issue.

6. Conclusion

Although the literature on solar radiation modeling is quite abundant, methodological developments on the topic of model validation and performance assessment have not improved much until recently. The metrics of model performance that are used in most of the current literature are still essentially the same as four decades ago. This investigation has underlined the importance of appropriate validation of the solar radiation datasets and their underlying models that are routinely used in solar resource assessment for the design and financing of large solar projects. Higher solar radiation data accuracy translates into lower financial risks and better bankability.

Conducting validation studies of solar radiation models and data can be done in different ways, depending on the goal of the study and, most importantly, on the degree of certainty of the inputs to the model under scrutiny. When these inputs are highly accurate, it is possible to validate the model itself. In most practical situations, however, this is not the case and only the combination *model+inputs* can then be validated.

Another important issue when dealing with any type of validation is how to correctly evaluate the performance of a model or its predictions. This contribution examines different metrics, and separates them into four different classes. Some of these metrics (those in Class A) are quite conventional and most certainly well-known from the whole solar community. Those in Class B are performance indicators that attempt to combine the merits of some of the Class-A statistics. They are not as commonplace in the literature, however. Metrics of Class C are more elaborate because they compare the shapes of the modeled and reference frequency distributions to evaluate their likeliness. A correct frequency distribution of the incident irradiance is necessary, for instance, to lower the uncertainty in the predicted power output of non-linear systems, such as concentrating solar power plants. The recently introduced CPI metric combines the advantages of some indicators from Class A and C, and is shown to be much more discriminant than other statistics when comparing different models of relatively comparable performance. Finally, visual indicators of Class D provide additional information that cannot be described by a single statistic.

Using high-quality experimental measurements of direct normal irradiance (DNI) at Tamanrasset, Algeria, a practical example is developed to obtain all the different indicators reviewed here. A visual analysis of the response of DNI to the jointly measured aerosol optical depth (AOD) shows that this is a critical aspect of solar radiation modeling, particularly over areas with occasional or frequent high-turbidity conditions. An important finding is that some models that return good predictions under low-AOD conditions may become unacceptable beyond a critical AOD threshold. Unfortunately, the limits of validity of the radiation models currently in use to develop solar resource data, or other applications, are not always mentioned by model authors. Even if they are, their performance beyond these limits may or may not be acceptable. This issue should be taken into consideration by all providers of solar resource data.

A desirable outcome of this study would be the development of international guidelines about how to conduct validation studies, and what specific metrics to use in order to provide the best information on the accuracy of modeled solar radiation data, which could be used by modelers to improve their models, and by non-technical stakeholders to accelerate their financial analyses and ultimately improve the bankability of large-scale solar projects.

Acknowledgments

The AERONET and WRMC-BSRN staff and participants are thanked for their successful effort in establishing and maintaining the Tamanrasset site whose data were advantageously used in this investigation. The author is also grateful to Dr. Jesus Polo and Dr. Jose-Antonio Ruiz-Arias for their help with the development of customized codes for the preparation of KS-related metrics and Taylor diagrams, respectively. Finally, the author acknowledges the partial financial support provided by the National Renewable Energy Laboratory through subcontract AGG-3-23359-01, and by the University Corporation for Atmospheric Research through subaward Z13-13584.

References

[1] Cho H, Fumo N. Uncertainty analysis for dimensioning solar photovoltaic arrays. In: Proceedings of the 6th international conference on energy sustainability. San Diego, CA: ASME; 2012.

[2] Ho CK, Khalsa SS, Kolb GJ. Methods for probabilistic modeling of concentrating solar power plants. Sol Energy 2011;85:669–75.

[3] Leloux J, Lorenzo E, Garcia B, Aguilera J, Gueymard CA. A bankable method for the field testing of a CPV plant. Appl Energy 2014;118:1–11.

[4] Price DW, Goedeke SM, Lausten MW, Kirkpatrick K. Understanding solar power performance risk and uncertainty. In: Proceedings of the 6th International conference on energy sustainability. San Diego, CA: ASME; 2012.

[5] Davies JA, McKay DC. Estimating solar irradiance and components. Sol Energy 1982;29:55–64.

[6] Hoff TE, Perez P. Predicting short-term variability of high-penetration PV. In: World renewable energy forum. Denver, CO: American Solar Energy Society; 2012.

[7] Cebecauer T, Suri M, Gueymard CA. Uncertainty sources in satellite-derived direct normal irradiance: how can prediction accuracy be improved globally? In: Proceedings of the SolarPACES conference. Granada, Spain; 2011. <http://www.solarconsultingservices.com/Cebecauer-Uncertainty%20satell%20data-SolarPACES11.pdf>.

[8] Vignola F, Grover C, Lemon N, McMahan A. Building a bankable solar radiation dataset. Sol Energy 2012;86:2218–29.

[9] Schnitzer M, Thuman C, Johnson P. The impact of solar uncertainty on project financeability: mitigating energy risk through on-site monitoring. In: World renewable energy forum. Denver, CO: American Solar Energy Society; 2012.

[10] Suri M, Cebecauer T, Perez P. Quality procedures of SolarGIS for provision site-specific solar resource information. In: Proceedings of the SolarPACES conference. Perpignan, France; 2010.

[11] Gueymard CA, Gustafson B, Bender G, Etringer A, Storck P. Evaluation of procedures to improve solar resource assessments: optimum use of short-term data from a local weather station to correct bias in long-term satellite derived solar radiation time series. In: World renewable energy forum. Denver, CO: American Solar Energy Society; 2012. <http://www.solarconsultingser

vices.com/Gueymard&3Tier-Procedures%20to%20Improve%20Solar%20Resource%20Assessments-WREF12.pdf > .

[12] Thuman C, Schertzer W, Johnson P. Quantifying the accuracy of the use of Measure–Correlate–Predict methodology for long-term solar resource estimates. In: World renewable energy forum. Denver, CO: American Solar Energy Society; 2012.

[13] Meyer R, Butron JT, Marquardt G, Schwandt M, Geuder N, Hoyer-Klick C, et al. Combining solar irradiance measurements and various satellite-derived products to a site-specific best estimate. In: Proceedings of the SolarPACES Conference. Las Vegas, NV; 2008.

[14] Renné D, Beyer HG, Wald L, Meyer R, Perez P, Stackhouse P. Solar resource knowledge management: a new task of the International Energy Agency. In: Proceedings of the solar conference. Denver, CO: American Solar Energy Society; 2006.

[15] Renné D. Status of Task 36 solar resource knowledge management under the IEA Solar Heating and Cooling program. In: Proceedings of the solar conference. Buffalo, NY: American Solar Energy Society; 2009.

[16] Beyer H-G, Polo J, Suri M, Torres JL, Lorenz E, Muller SC, et al. Report on benchmarking of radiation products. MESoR 2009. 〈http://www.mesor.org/docs/MESoR_Benchmarking_of_radiation_products.pdf〉.

[17] Dumortier D. Description of solar resource products, summary of benchmarking results, examples of use. MESoR 2009. 〈http://www.mesor.org/docs/MESoR_Handbook_on_best_practices.pdf〉.

[18] Hoyer-Klick C, Beyer H-G, Dumortier D, Schroedter-Homscheid M, Wald L, Martinoli M, et al. MESoR—Management and Exploitation of Solar Resource Knowledge. In: Proceedings of the SolarPACES conference. Berlin, Germany; 2009.

[19] SolarPACES. Annual report, part 6. 〈http://www.solarpaces.org/Tasks/Task5/documents/Part6_SolarPACES2011_AnnualReport_Part6_SolRes_20120806-2 TaskV.pdf〉; 2011.

[20] Meyer R, Gueymard CA, Ineichen P. Standardizing and benchmarking of modeled DNI data products. In: Proceedings of the SolarPACES conference. Granada, Spain; 2011. < http://www.solarconsultingservices.com/DNI%20modeled%20data%20benchmarking-SolarPACES11.pdf > .

[21] Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim Res 2005;30:79–82.

[22] Willmott CJ, Matsuura K, Robeson SM. Ambiguities inherent in sums-of-squares-based error statistics. Atmos Environ 2009;43:749–52.

[23] Willmott CJ. On the validation of models. Phys Geogr 1981;2:184–94.

[24] Willmott CJ. Some comments on the evaluation of model performance. Bull Am Meteorol Soc 1982;63:1309–13.

[25] Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, et al. Statistics for the evaluation and comparison of models. J Geophys Res 1985;90C:8995–9005.

[26] Power HC. Estimating clear-sky beam irradiation from sunshine duration. Sol Energy 2001;71:217–24.

[27] Willmott CJ, Robeson SM, Matsuura K. A refined index of model performance. Int J Climatol 2012:32.

[28] Legates DR, McCabe GJ. A refined index of model performance: a rejoinder. Int J Climatol 2013;33:1053–6.

[29] Legates DR, McCabe GJ. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resour Res 1999;35:233–41.

[30] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models, I. A discussion of principles. J Hydrol 1970;10:282–90.

[31] Stone RJ. Improved statistical procedure for the evaluation of solar radiation estimation models. Sol Energy 1993;51:289–91.

[32] Stone RJ. A nonparametric statistical procedure for ranking the overall performance of solar radiation models at multiple locations. Energy 1994;19:765–9.

[33] Muneer T, Younes S, Munawwar S. Discourses on solar radiation modeling. Renew Sustain Energy Rev 2007;11:551–602.

[34] Gueymard CA, Myers DR. Validation and ranking methodologies for solar radiation models. In: Badescu V, editor. Modeling solar radiation at the earth surface. Springer; 2008.

[35] Badescu V. Assessing the performance of solar radiation computing models and model selection procedures. J Atmos Sol-Terr Phys 2013;105:119–34.

[36] Hoff TE, Perez R, Kleissl J, Renné D, Stein J. Reporting of irradiance modeling relative prediction errors. Prog Photovolt: Res Appl 2012;21:1514–9.

[37] Moriasi DN, Arnold JG, Vaan Liew MW, Bingner RL, Harmel RD, Veith TL. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans Am Soc Agric Biol Eng 2007;50:885–900.

[38] Meyer R, Beyer HG, Fanslau J, Geuder N, Hammer A, Hirsch T, et al. Towards standardization of CSP yield assessments. In: Proceedings of the SolarPACES conference. Berlin, Germany; 2009.

[39] Marquez R. Coimbra CFM. Proposed metric for evaluation of solar forecasting models. J SOL ENERG-T ASME 2012;135:011016. http://dx.doi.org/10.1115/1.4007496.

[40] Oreskes N, Shrader-Frechette K, Belitz K. Verfication, validation, and confirmation of numerical models in the Earth sciences. Science 1994;263:641–6.

[41] Gubler S, Gruber S, Purves RS. Uncertainties of parameterized surface downward clear-sky shortwave and all-sky longwave radiation. Atmos Chem Phys 2012;12:5077–98.

[42] Gueymard CA. Direct solar transmittance and irradiance predictions with broadband models. Pt 2: Validation with high-quality measurements. Sol Energy 2003;74:381–95;
Gueymard CA. Corrigendum: Sol Energy 2004;76:515.

[43] Gueymard CA. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. Sol Energy 2012;86:2145–69.

[44] Halthore RN, Crisp D, Schwartz SE, Anderson GP, Berk A, Bonnel B, et al. Intercomparison of shortwave radiative transfer codes and measurements. J Geophys Res 2005:110D. http://dx.doi.org/10.1029/2004JD005293.

[45] Louche A, Simonnot G, Iqbal M. Experimental validation of some clear-sky insolation models. Sol Energy 1988;41:273–9.

[46] Antonanzas-Torres F, Canizares F, Perpinan O. Comparative assessment of global irradiation from a satellite estimate model (CM SAF) and on-ground measurements (SIAR): A Spanish case study. Renew Sustain Energy Rev 2013;21:248–61.

[47] Badescu V, Dumitrescu A. The CMSAF hourly solar irradiance database (product CM54): accuracy and bias corrections with illustrations for Romania (south-eastern Europe). J Atmos Ocean Technol 2013;93:100–9.

[48] Badescu V, Gueymard CA, Cheval S, Oprea C, Baciu M, Dumitrescu A, et al. Computing global and diffuse solar hourly irradiation on clear sky. Review and testing of 54 models. Renew Sustain Energy Rev 2012;16:1636–56.

[49] Battles FJ, Olmo FJ, Tovar J, Alados-Arboledas L. Comparison of cloudless sky parameterizations of solar irradiance at various Spanish midlatitude locations. Theor Appl Climatol 2000;66:81–93.

[50] Huang G, Ma M, Liang S, Liu S, Li X. A LUT—based approach to estimate surface solar irradiance by combining MODIS and MTSAT data. J Geophys Res 2011:116D. http://dx.doi.org/10.1029/2011JD016120.

[51] Ineichen P, Barroso CS, Geiger B, Hollman R, Marsouin A, Mueller R. Satellite application facilities irradiance products: hourly time step comparison and validation over Europe. Int J Remote Sens 2009;30:5549–71.

[52] Ineichen P. Five satellite products deriving beam and global irradiance validation on data from 23 ground stations. Switzerland: University of Geneva; 2011.

[53] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos Meas Tech 2013;6:2403–18.

[54] Li Z, Whitlock CH. Charlock TP. Assessment of the global monthly mean insolation estimated from satellite measurements using global energy balance archive data. J Clim 1995;8:315–28.

[55] Longman R, Giambelluca TW, Frazier AG. Modeling clear-sky solar radiation across a range of elevations in Hawai'i: comparing the use of input parameters at different temporal resolutions. J Geophys Res 2012:117D. http://dx.doi.org/10.1029/2011JD016388.

[56] Nemes C. A clear sky irradiation assessment using the European Solar Radiation Atlas model and Shuttle Radar Topography Mission database: a case study for Romanian territory. J Renew Sustain Energy 2013;5:041807.

[57] Nonnenmacher L, Kaur A. Coimbra CFM. Verification of the SUNY direct normal irradiance model with ground measurements. Sol Energy 2014;99: 246–258.

[58] Pagola I, Gaston M, Fernandez-Peruchena C, Moreno S, Ramirez L. New methodology of solar radiation evaluation using free access databases in specific locations. Renew Energy 2010;35:2792–8.

[59] Perez P, Ineichen P, Moore K, Kmiecik M, Chain C, George R, et al. A new operational model for satellite-derived irradiances: description and validation. Sol Energy 2002;73:307–17.

[60] Polo J, Zarzalejo LF, Cony M, Navarro AA, Marchante R, Martin L, et al. Solar radiation estimations over India using Meteosat satellite images. Sol Energy 2011;85:2395–406.

[61] Posselt R, Mueller RW, Stöckli R, Trentmann J. Remote sensing of solar surface radiation for climate monitoring—the CM-SAF retrieval in international comparison. Rem Sens Environ 2012;118:186–98.

[62] Schillings C, Meyer R, Mannstein H. Validation of a method for deriving high resolution direct normal irradiance from satellite data and application for the Arabian Peninsula. Sol Energy 2004;76:485–97.

[63] Schroeder TA, Hember R, Coops NC. Validation of solar radiation surfaces from MODIS and reanalysis data over topographically complex terrain. J Appl Meteorol Climatol 2009;48:2441–58.

[64] Zhang T, Stackhouse PW, Gupta SK, Cox SJ, Mikovitz JC, Hinkelman L. The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: a systematic quality control, production and application approach. J Quant Spectrosc Radiat Transf 2013;122:127–40.

[65] Elagib NA, Mansell MG. New approaches for estimating global solar radiation across Sudan. Energy Convers Manag 2000;41:419–34.

[66] Wilcox S. National solar radiation database 1991–2010 Update: User's manual. Golden, CO: National Renewable Energy Laboratory; 2012.

[67] Long CN. The Shortwave (SW) Clear-Sky Detection and Fitting Algorithm: Algorithm Operational Details and Explanations. Atmos Radiat Meas Prog, Rep. ARM TR-004, DOE 2000.

[68] Long CN, Ackerman TP. Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. J Geophys Res 2000;105D:15609–26.

[69] Gueymard CA, Wilcox SM. Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data. Sol Energy 2011;85:1068–84.

[70] Roesch A, Wild M, Ohmura A, Dutton EG, Long CN, Zhang T. Assessment of BSRN radiation records for the computation of monthly means. Atmos Meas Tech 2011;4:339–54 (Corrigendum) http//www.atmos-meas-tech.net/4/973/2011/.

[71] Polo J, Zarzalejo LF, Ramirez L, Espinar B. Iterative filtering of ground data for qualifying statistical models for solar irradiance estimation from satellite data. Sol Energy 2006;80:240–7.
[72] Espinar B, Ramirez L, Drews A, Beyer HG, Zarzalejo LF, Polo J, et al. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Sol Energy 2009;83:118–25.
[73] Marsaglia G, Tsang WW, Wang J. Evaluating Kolmogorov's Distribution. J Stat Softw 2003;8:1–4.
[74] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 2001;106D:7183–92.
[75] Correa CD, Lindstrom P. The mutual information diagram for uncertainty visualization. Int J Uncertain Quantif 2013;3:187–201.
[76] Ineichen P., Guisan O, Razafindraibe A. Indice de clarté. Mesures d'ensoleillement à Genève, vol 9. Available from ⟨http://wwwcuepech/html/biblio/pdf/ineichen%201984%20-%20indice%20de%20clarte%20(9)pdf⟩: Switzerland: University of Geneva; 1984.
[77] Ineichen P, Perez P, Seals R. The importance of correct albedo determination for adequately modeling energy received by tilted surfaces. Sol Energy 1987;39:301–5.
[78] Gueymard CA, Myers DR. Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. Sol Energy 2009;83:171–85.
[79] Davies JA, McKay DC. Evaluation of selected models for estimating solar radiation on horizontal surfaces. Sol Energy 1989;43:153–68.
[80] Meinel AB, Meinel MP. Applied solar energy, an introduction. Reading, MA, USA: Addison-Wesley; 1976.
[81] Renno C, Petito F. Design and modeling of a concentrating photovoltaic thermal (CPV/T) system for a domestic application. Energy Build 2013;62:392–402.
[82] Allen RG. Assessing integrity of weather data for use in reference evapotranspiration estimation. J Irrigat Drain Eng, Am Soc Civ Eng 1996;122:97–106.
[83] Allen RG, Trezz R, Tasumi M. Analytical integrated functions for daily solar radiation on slopes. Agric Meteorol 2006;139:55–73.
[84] ASCE-EWRI. The ASCE standardized reference evapotranspiration equation. Washington, DC: Environmental and Water Resources Institute (EWRI) of the American Society of Civil Engineers (ASCE) task committee on standardization of reference evapotranspiration calculation; available from ⟨http://www.kimberly.uidaho.edu/water/asceewri/⟩; 2005.
[85] Majumdar NC, Mathur BL, Kaushik SB. Prediction of direct solar radiation for low atmospheric turbidity. Solar Energy 1972;13:383–94.
[86] Boes EC. Fundamentals of solar radiation. In: Kreider JF, Kreith F, editors. Solar energy handbook. New York: McGraw Hill; 1981.
[87] Ineichen P, Perez P. A new airmass independent formulation for the Linke turbidity coefficient. Sol Energy 2002;73:151–7.
[88] Ineichen P. A broadband simplified version of the Solis clear sky model. Sol Energy 2008;82:758–62.
[89] Bird RE, Hulstrom RL. A simplified clear sky model for direct and diffuse insolation on horizontal surfaces. Golden, CO: Solar Energy Research Institute (now NREL); 1981. ⟨http://www.nrel.gov/docs/legosti/old/761.pdf⟩.
[90] Bird RE, Hulstrom RL. Review, evaluation, and improvement of direct irradiance models. J SOL ENERG-T ASME 1981;103:182–92.
[91] Gueymard CA. Direct solar transmittance and irradiance predictions with broadband models. Pt 1: Detailed theoretical performance assessment. Sol Energy 2003;74:355–79;
Gueymard CA. Corrigendum: Sol Energy 2004;76:513.
[92] Gueymard CA. REST2: High performance solar radiation model for cloudless-sky irradiance, illuminance and photosynthetically active radiation—validation with a benchmark dataset. Sol Energy 2008;82:272–85.
[93] Gueymard CA. Temporal variability in direct and global irradiance at various time scales as affected by aerosols. Sol Energy 2012;86:3544–53.
[94] Vindel JM, Polo J, Antonanzas-Torres F. Improving daily output of global to direct solar irradiance models with ground measurements. J Renew Sustain Energy 2014;5:063123.